AD-A057 873 SYRACUSE UNIV N Y F/G 9/2

BAYESIAN SOFTWARE PREDICTION MODELS. VOLUME IV. BAYESIAN SOFTWA--ETC(U)

JUL 78 A L GOEL, K OKUMOTO F30602-76-C-0097

UNCLASSIFIED TR-78-4 RADC-TR-78-155-VOL-4 NL

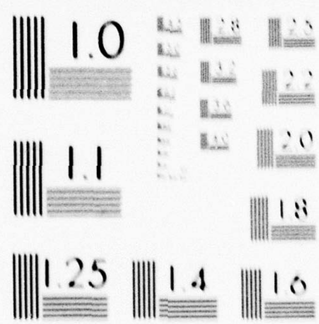
1 OF 1

AD
A057 873

END

DATE
FILMED
10-78

DDC



MICROCOPY RESOLUTION TEST CHART

ADA057873

AD No. ____
DDC FILE COPY

(2) LEVEL III

A057872
A057870
A057871

RADC-TR-78-155, Volume IV (of five)
Final Technical Report
July 1978

BAYESIAN SOFTWARE PREDICTION MODELS
Bayesian Software Correction Limit Policies

Amrit L. Goel
K. Okumoto

Syracuse University

Approved for public release; distribution unlimited.

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

78 08 22 063

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-78-155, Volume IV (of five) has been reviewed and is approved for publication.

APPROVED: Alan N. Sukert

ALAN N. SUKERT
Project Engineer

APPROVED: Alan R. Barnum

ALAN R. BARNUM
Assistant Chief
Information Sciences Division

FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

(19) TR-78-155-VOL-4 (14) TR-78-4

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

(18) 1. REPORT NUMBER: RADC-TR-78-155, Vol IV (of five)

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER (9)

(6) 4. TITLE (and Subtitle): BAYESIAN SOFTWARE PREDICTION MODELS. Volume IV. Bayesian Software Correction Limit Policies.

5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report, Dec 75 - Mar 78

6. PERFORMING ORG. REPORT NUMBER: Technical Report No. 78-4

(10) 7. AUTHOR(s): Amrit L. Goel, K. Okumoto

(15) 8. CONTRACT OR GRANT NUMBER(s): F30602-76-C-0097

See reverse

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Syracuse University, Syracuse NY 13210

(16) 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 62702F, 55811403 (17) 14

11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (ISIS), Griffiss AFB NY 13441

(11) 12. REPORT DATE: July [redacted]978

13. NUMBER OF PAGES: 36

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Same

15. SECURITY CLASS. (of this report): UNCLASSIFIED (12) 41p.

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DDC
RECEIVED
AUG 23 1978
B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer: Alan N. Sukert (ISIS)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Software Error Correction
Bayesian Estimates
Software Correction Costs

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report deals with the problem of determining an optimum correction limit policy for a large software system subject to random occurrences of errors in an operational phase. When an error occurs, the corrective action is scheduled for either the programmer (Phase I) or the system analyst (Phase II). Two cost models are developed and procedures for obtaining the optimum correction time T* are described to minimize the expected cost per unit time. Numerical examples are given to illustrate the procedure.

(Cont'd)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

339 600

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

Volume V will be published at a later date.

Block 7.

[1]Professor, Department of Industrial Engineering & Operations Research, and School of Computer and Information Science, Syracuse University.

[2]Research Assistant, Department of Industrial Engineering & Operations Research, Syracuse University.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EVALUATION

The necessity for more complex software systems in such areas as command and control and avionics has led to the desire for better methods for predicting software errors to insure that software produced is of higher quality and of lower cost. This desire has been expressed in numerous industry and Government sponsored conferences, as well as in documents such as the Joint Commanders' Software Reliability Working Group Report (Nov 1975). As a result, numerous efforts have been initiated to develop methods for determining the optimal policy for maintenance of an operational software system. However, early efforts have not developed any consistent or generally applicable software maintenance policy.

This effort was initiated in response to this need for developing better and more accurate software error prediction models and fits into the goals of RADC TPO No. 5, Software Cost Reduction (formerly RADC TPO No. II, Software Sciences Technology), in the subthrust of Software Quality (Software Modeling). This report summarizes the development of a Bayesian methodology for determining the optimal policy for maintaining an operational software system. The importance of this development is that it represents the first attempt to develop operational software maintenance policies that more closely reflect the actual software error detection and correction process.

The theory and equations developed under this effort will lead to much needed predictive measures for use by software maintenance personnel in providing better and more efficient maintenance of operational software. In addition, the associated confidence limits and other related statistical quantities developed under this effort will insure more widespread use of these modeling techniques. Finally, the predictive measures and equations developed under this effort will be applicable to current Air Force software development projects and thus help to produce the high quality, low cost software needed for today's systems.

Alan N. Sukert

ALAN N. SUKERT
Project Engineer

1. INTRODUCTION

In this report we discuss the problem of determining an optimum correction limit policy for a large software system which is subject to random occurrences of errors. When an error occurs, a corrective action is undertaken to remove it. Such an action can be scheduled at two levels, which we call Phase I and Phase II. By Phase I we mean that the corrective action will be undertaken by the programmer while Phase II action is undertaken by a system analyst or system designer. First, Phase I corrective action is scheduled for a specified time T. If the error is not corrected in this time, it is referred to Phase II. This sequence of corrective actions in an operational phase is shown in Figure 1.1. Our objective is to determine the optimum value T^* of T which minimizes the long run average cost. Two models are developed for this purpose. In the first model (Section 2) we assume that the cost of observations of error occurrence and correction time, prior to the implementation of the optimum policy, is negligible. The second model (Section 3) incorporates the cost of observations.

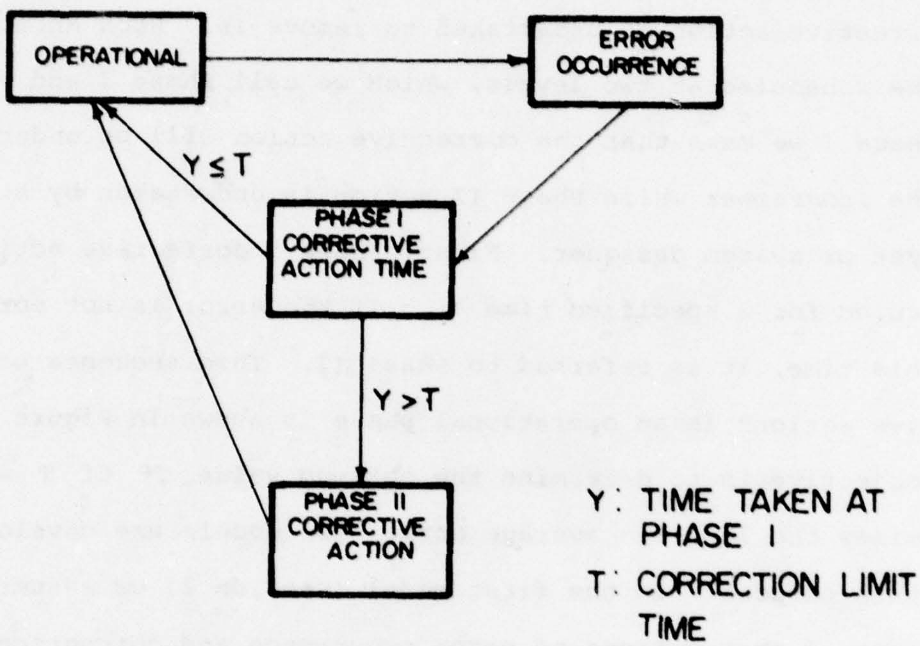


Figure 1.1 Sequence of Corrective Actions in Operational Phase

2. MODEL WHEN COST OF OBSERVATIONS IS NEGLIGIBLE

The following assumptions are made for model development:

(i) The error occurrence time in a software system has an exponential distribution with an unknown mean λ.

(ii) The error correction time at Phase I is exponential with an unknown mean μ_1.

(iii) The Phase II error correction time has a general distribution with a known mean μ_2.

(iv) Appropriate prior distributions can be chosen for λ and μ_1.

2.1 Predictive Distributions of Y and X

Let random variables X and Y denote the error occurrence time and Phase I error correction time, respectively. From assumptions (i) and (ii) the probability density functions are

$$f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda} \qquad x > 0, \quad \lambda > 0 \tag{2.1}$$

$$g(y|\mu_1) = \frac{1}{\mu_1} e^{-y/\mu_1} \qquad y > 0, \quad \mu_1 > 0 \tag{2.2}$$

In this report, we develop suitable expressions by considering the conjugate priors for μ_1 and λ which are inverted gamma distributions given by

$$P(\mu_1) = \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \mu_1^{-(\alpha_1+1)} e^{-\beta_1/\mu_1} \qquad \alpha_1, \beta_1 > 0, \tag{2.3}$$

and

$$P(\lambda) = \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} \lambda^{-(\alpha_2+1)} e^{-\beta_2/\lambda} \qquad \alpha_2, \beta_2 > 0. \tag{2.4}$$

The expressions for any other reasonable priors for μ_1 and λ can be developed similarly.

Also, let $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ and $\underset{\sim}{y} = (y_1, y_2, \ldots, y_n)$ be the observed values of n error occurrence times and n error correction times, respectively.

Now we obtain expressions for the predictive distributions of Y and X and also obtain the Bayesian estimates $\hat{y}_{n+1}$ and $\hat{x}_{n+1}$ which will be used to obtain the cost function.

For given observations $\underset{\sim}{y}$, the likelihood function of μ_1 is

$$\ell(\mu_1 | \underset{\sim}{y}) = \mu_1^{-n} e^{-\sum_{i=1}^{n} y_i/\mu_1} . \tag{2.5}$$

The posterior distribution of μ_1 is obtained from Bayes theorem:

$$P(\mu_1 | \underset{\sim}{y}) = \frac{\ell(\mu_1 | \underset{\sim}{y}) P(\mu_1)}{\int \ell(\mu_1 | \underset{\sim}{y}) P(\mu_1) d\underset{\sim}{y}} . \tag{2.6}$$

Substituting the expressions for $p(\mu_1)$ and $\ell(\mu_1 | y)$ from (2.3) and (2.5), we get the posterior distribution of μ_1 as

$$P(\mu_1 | \underset{\sim}{y}) = \frac{(\beta_1 + \sum_{i=1}^{n} y_i)^{\alpha_1 + n}}{\Gamma(\alpha_1 + n)} \cdot \mu_1^{-(\alpha_1 + n + 1)} e^{-\{(\beta_1 + \sum_{i=1}^{n} y_i)/\mu_1\}} \tag{2.7}$$

Using this posterior, the predictive distribution $g(y | \underset{\sim}{y})$ of error correction time at Phase I is given by

$$g(y | \underset{\sim}{y}) = \int_0^\infty g(y | \mu_1) P(\mu_1 | \underset{\sim}{y}) d\mu_1 . \tag{2.8}$$

Substituting the expressions for $g(y|\mu_1)$ and $p(\mu_1|\underset{\sim}{y})$ from (2.2) and (2.6), respectively, we get

$$g(y|\underset{\sim}{y}) = \left(\frac{n+\alpha_1}{\sum_{i=1}^{n} y_i + \beta_1}\right)\left(1 + \frac{y}{\sum_{i=1}^{n} y_i + \beta_1}\right)^{-(n+\alpha_1+1)} . \tag{2.9}$$

The cumulative predictive distribution to some specified time t is

$$G(t|\underset{\sim}{y}) = \int_0^t g(y|\underset{\sim}{y})dy$$

$$= 1 - \left(\frac{t}{\sum_{i=1}^{n} y_i + \beta_1}\right)^{-(n+\alpha_1)} . \tag{2.10}$$

We define the predictive Phase I error correction rate as

$$r(t|\underset{\sim}{y}) = \frac{g(t|\underset{\sim}{y})}{\bar{G}(t|\underset{\sim}{y})} \tag{2.11}$$

so that

$$r(t|\underset{\sim}{y}) = \left(\frac{n+\alpha_1}{\sum_{i=1}^{n} y_i + \beta_1}\right)\left(1 + \frac{t}{\sum_{i=1}^{n} y_i + \beta_1}\right)^{-1} \tag{2.12}$$

where $\bar{G}(t) \equiv 1 - G(t)$.

The predictive distribution $f(x|\underset{\sim}{x})$ of the error occurrence time can be similarly obtained.

2.2 Bayesian estimates $\hat{x}_{n+1}$ and $\hat{y}_{n+1}$.

From the predictive distributions of X and Y the Bayesian estimates $\hat{x}_{n+1}$ of the time to (n+1)st error occurrence, for given $\underset{\sim}{x}$ and the (n+1)st error correction time for given $\underset{\sim}{y}$ are easily obtained, since

$$\hat{y}_{n+1} = \int_0^\infty \bar{G}(t|\underset{\sim}{y})dt$$

$$= \frac{\sum_{i=1}^{n} y_i + \beta_1}{\alpha_1 + n-1} \tag{2.13}$$

Similarly

$$\hat{x}_{n+1} = \int_0^\infty \bar{F}(t|\underset{\sim}{x})dt , \tag{2.14}$$

where

$$F(t|\underset{\sim}{x}) = \int_0^t f(x|\underset{\sim}{x})dx \tag{2.15}$$

which is a cumulative predictive distribution to some specified time t. Hence

$$\hat{x}_{n+1} = \frac{\sum_{i=1}^{n} x_i + \beta_2}{\alpha_2 + n-1} . \tag{2.16}$$

2.3 Cost Function

Let c_1 (c_2) be the cost per unit time of error correction in Phase I (Phase II) and the costs be linear functions of time. From assumption (iii) Phase II error correction time has some arbitrary general distribution with a known mean μ_2. If we consider one cycle to be the time from the beginning of (n+1)st operation to the beginning of (n+2)nd operation, then the expected cost in one cycle is

$$E(C) = c_1 \int_0^T \bar{G}(t|\underset{\sim}{y})dt + c_2\mu_2\bar{G}(T|\underset{\sim}{y}) , \tag{2.17}$$

where T denotes the scheduled correction limit time in Phase I.

The expected length of one cycle is

$$E(L) = \hat{x}_{n+1} + \int_0^T \overline{G}(t|\underline{y})dt + \mu_2 \overline{G}(T|\underline{y}) \ . \tag{2.18}$$

and hence the long run expected cost per unit time is

$$C(T) = \frac{E(C)}{E(L)}$$

or

$$C(T) = \frac{c_1 \int_0^T \overline{G}(t|\underline{y})dt + c_2 \mu_2 \overline{G}(T|\underline{y})}{\hat{x}_{n+1} + \int_0^T \overline{G}(t|\underline{y})dt + \mu_2 \overline{G}(T|\underline{y})} \ . \tag{2.19}$$

This is the cost function which we want to optimize to obtain the optimum policy.

2.4 Optimum Policy

From (2.19), we note that

$$C(0) = \frac{c_2 \mu_2}{\hat{x}_{n+1} + \mu_2} \tag{2.20}$$

and

$$C(\infty) = \frac{c_1 \hat{y}_{n+1}}{\hat{x}_{n+1} + \hat{y}_{n+1}} \tag{2.21}$$

where $\hat{y}_{n+1}$ is the Bayesian estimate of y for given data $\underline{y}$. Also, note that $T = 0$ means that the errors are corrected only at Phase II while $T = \infty$ means that they are corrected at Phase I.

To obtain an optimum T^* which minimizes the long run average cost per unit time, $C(T)$, we need the following theorems and corollary. Theorems 2.1 and 2.2 are the special cases of the theorems proved in Appendices A and B respectively.

Theorem 2.1

Assume $c_1 < c_2$. Then under the following condition

$$r(0|\underline{y}) > \frac{c_1(\hat{x}_{n+1} + \mu_2) - c_2\mu_2}{c_2\hat{x}_{n+1}\mu_2} \tag{2.22}$$

there exists a finite and unique T^* which satisfies

$$r(T|\underline{y})\{c_2\hat{x}_{n+1} + (c_2-c_1)\int_0^T \bar{G}(t|\underline{y})dt\}$$

$$+ (c_2-c_1)\bar{G}(T|\underline{y}) = \frac{c_1\hat{x}_{n+1}}{\mu_2}. \tag{2.23}$$

Theorem 2.2

If the above conditions are satisfied then there also exists a finite and unique upper bound $\bar{T}(> T^*)$ such that

$$r(\bar{T}|\underline{y}) = \frac{c_1\hat{x}_{n+1}}{\mu_2\{c_2\hat{x}_{n+1} + (c_2-c_1)\hat{y}_{n+1}\}}. \tag{2.24}$$

This upper bound can be used to obtain an initial value for solving the nonlinear equations in T^*.

Corollary 2.1

If there exists an optimum T^*, then the associated cost function is given by

$$C(T^*) = \frac{c_1 - c_2\mu_2 r(T^*|\underline{y})}{1-\mu_2 r(T^*|\underline{y})}. \tag{2.25}$$

2.5 Numerical Example

We use simulated data in this example to illustrate the calculations and nature of various quantities in the determination of T^*.

Let

$$c_1 = 8000 \qquad c_2 = 9000$$

$$\alpha_1 = 0 \qquad \beta_1 = 0$$

$$\alpha_2 = 0 \qquad \beta_2 = 0$$

$$\mu_2 = 0.7$$

The simulated data (x_n, y_n) are given in Table 2.1. Suppose n=10 data points are available. The Bayesian estimates of x_{11} and y_{11} are obtained from (2.16) and (2.13) as $\hat{x}_{11}=59.60$ and $\hat{y}_{11}=0.78$, respectively. Such values for various n are given in Table 2.2. For the case $n = 10$ we see that the optimum correction limit time is $T^* = 0.90$ hours and the corresponding minimum cost rate is $C(T^*) = 99.44$ dollars/hour.

Thus, for this set of data, we will schedule corrective action in Phase I for 0.90 hours and if it cannot be completed in this time, the software system will be referred to the system analyst for corrective action.

TABLE 2.1

Simulated Values of x_n and y_n

n	x_n (Hrs.)	y_n (Hrs.)	n	x_n (Hrs.)	y_n (Hrs.)
1	61.34	1.90	11	53.44	1.03
2	27.84	1.08	12	2.87	0.95
3	154.30	0.85	13	31.27	0.60
4	14.58	0.26	14	97.06	0.02
5	10.86	0.01	15	78.17	1.49
6	35.35	0.31	16	124.52	0.52
7	140.13	0.38	17	0.49	0.36
8	36.47	1.50	18	12.33	0.08
9	8.74	0.43	19	85.44	3.51
10	46.79	0.27	20	23.59	0.10

TABLE 2.2

Calculation for the Optimum Correction Time Policy

n	$\hat{x}_{n+1}$ (hr.)	$\hat{Y}_{n+1}$ (hr.)	T* (hr.)	C(T*)
2	89.17	2.98	0	70.10
3	121.74	1.92	0	51.45
4	86.02	1.36	0	72.65
5	67.23	1.02	0	92.74
6	60.85	0.88	0.32	101.05
7	74.07	0.80	0.73	80.11
8	68.70	0.90	0.02	90.78
9	61.20	0.84	0.38	100.60
10	59.60	0.78	0.90	99.44
11	58.98	0.80	0.66	102.87
12	53.88	0.82	0.50	113.82
13	52.00	0.80	0.68	116.85
14	55.46	0.74	1.45	103.80
15	57.09	0.79	0.75	106.51
16	61.58	0.77	1.02	97.47
17	57.76	0.75	1.45	101.22
18	55.09	0.71	2.16	101.14
19	56.78	0.86	0	109.61
20	55.03	0.82	0.13	112.97

2.6 Sensitivity Analysis of the Optimum Correction Time T^*.

To study the sensitivity of T^* to changes in various parameters, we look at the plot of T^* versus average correction time, $\bar{y} = \Sigma y_i/n$. Plots for various values of c_1, c_2, α_1, α_2, β_1, β_2, λ, μ_1 and μ_2 are given in Figures 2.1 to 2.4. In Figure 2.1, α_1 is varied while other parameters are kept constant. The effect of changing c_2 while keeping other factors constant is shown in Figure 2.2. Effects of changing μ_2 and n are shown in Figures 2.3 and 2.4, respectively. The following observations can be made from these figures.

(i) T^* increases with α_1 for fixed $\bar{y}$ and the slope of T^* vs $\bar{y}$ lines is independent of α_1 (Figure 2.1)

(ii) T^* increases with c_2 for fixed $\bar{y}$ and the slope of T^* vs $\bar{y}$ line is independent of c_2 (Figure 2.2)

(iii) T^* increases with μ_2 for fixed $\bar{y}$ and the slope of T^* vs $\bar{y}$ line is independent of μ_2 (Figure 2.3)

(iv) T^* increases with n for fixed $\bar{y}$ but the rate of increase decreases with $\bar{y}$ (Figure 2.4).

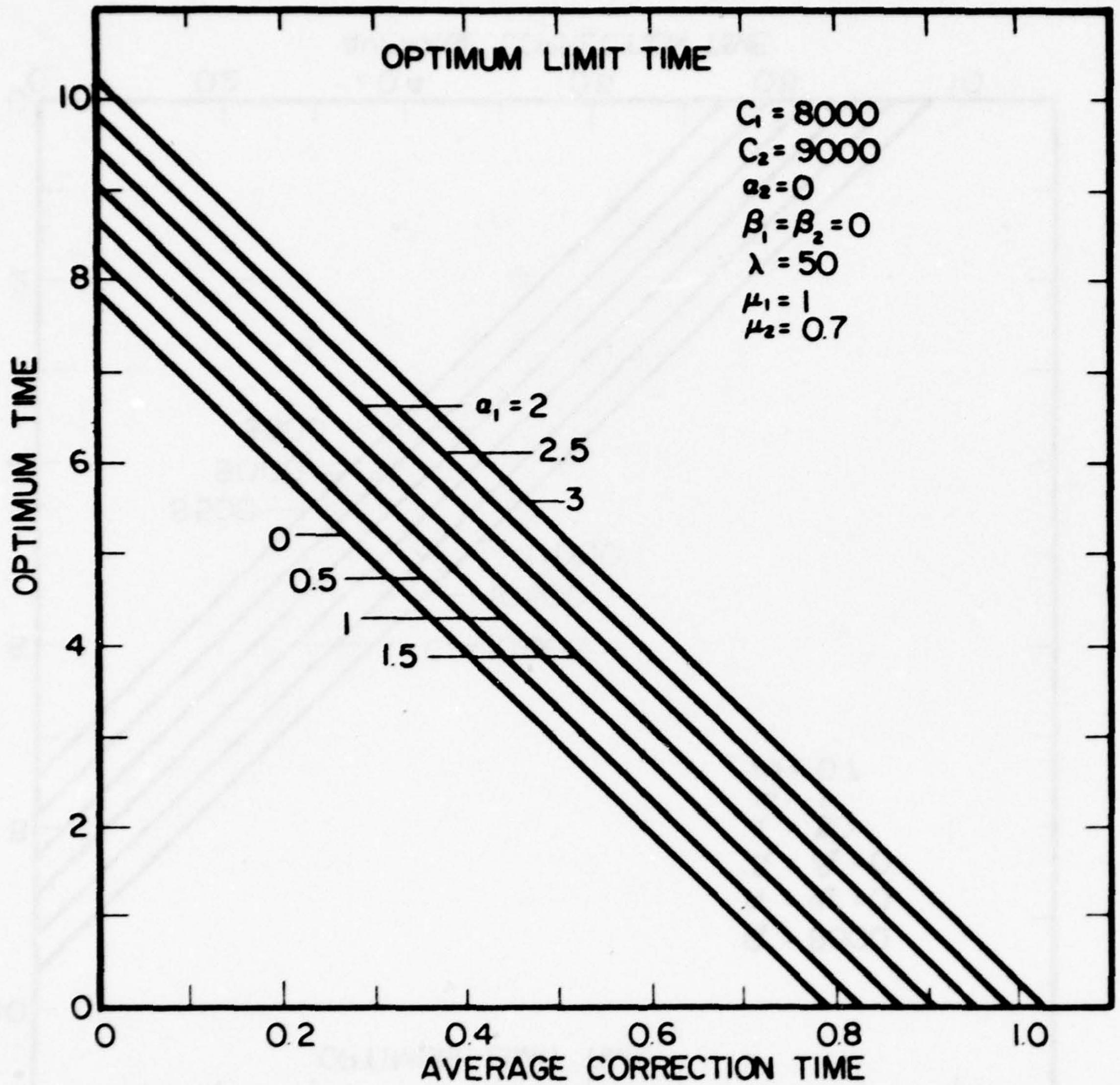


Figure 2.1 Plot of T^* vs $\bar{y}$ for Various Values of α_1

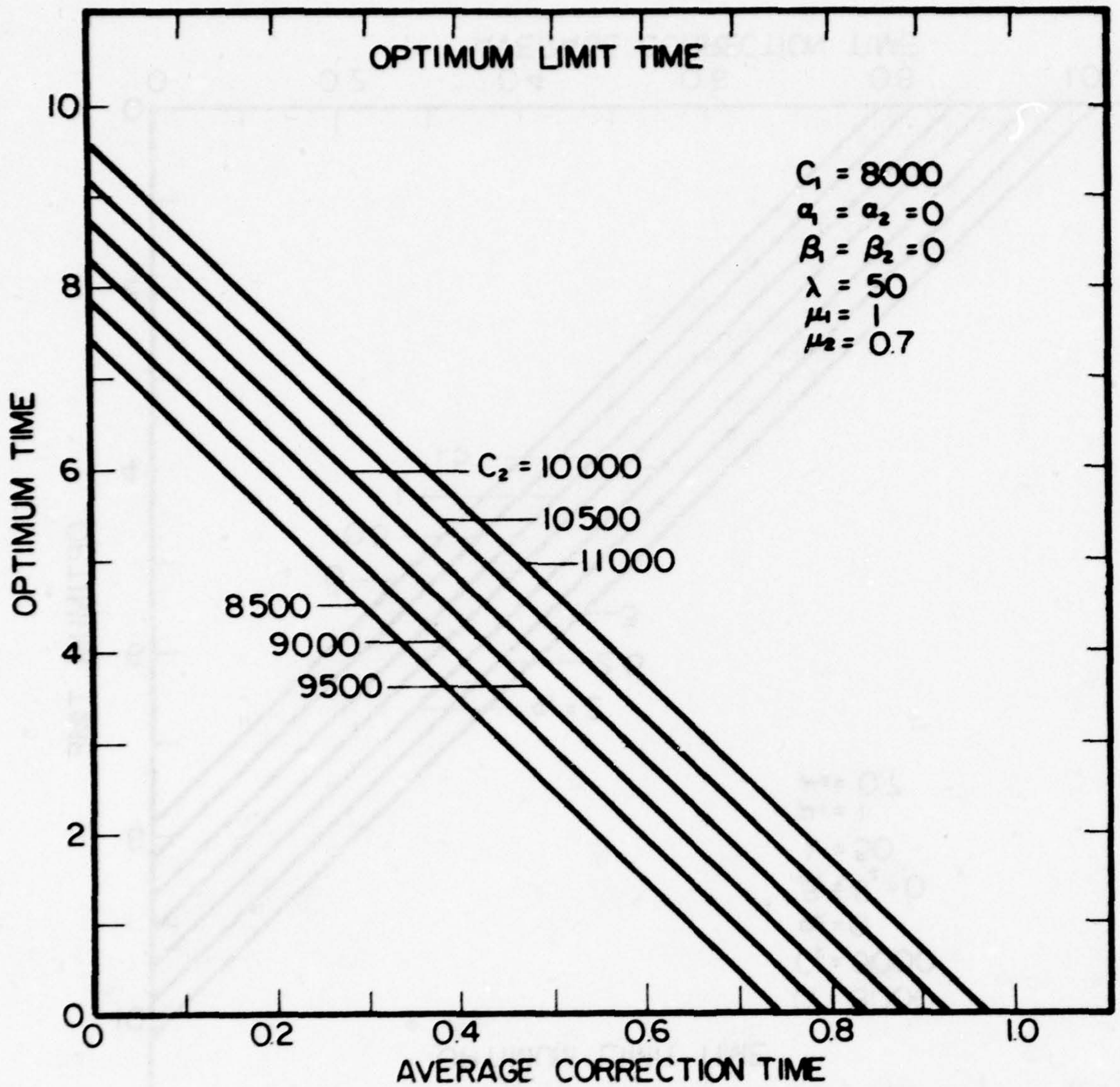


FIGURE 2.2 Plot of T* vs $\overline{y}$ for Various Values of c_2

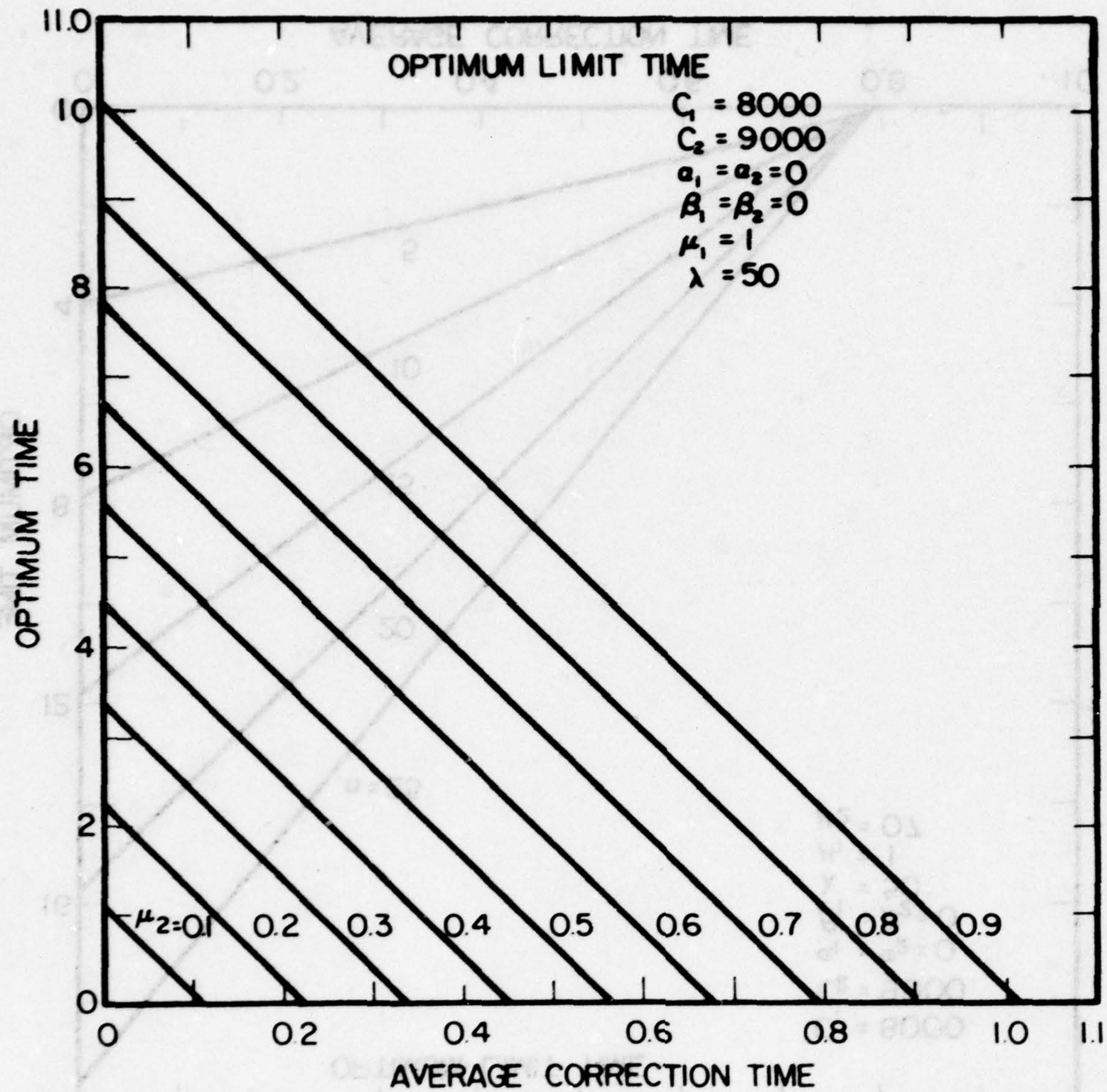


FIGURE 2.3 Plot of T* vs $\overline{y}$ for various Values of μ_2

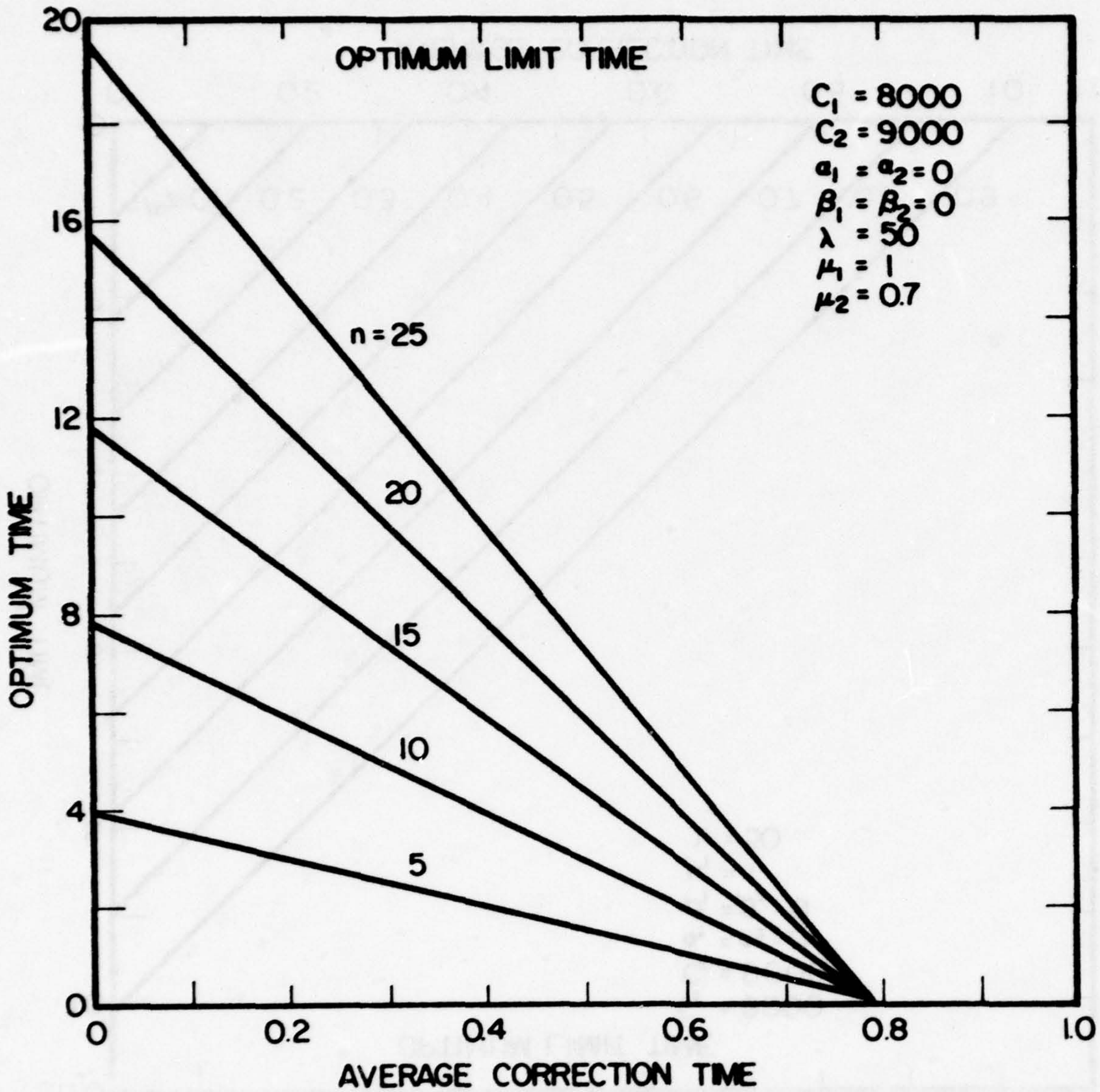


FIGURE 2.4 Plot of T^* vs $\bar{y}$ for Various Values of n

3. MODEL FOR NONZERO COST OF OBSERVATIONS

In this section we develop the cost model for the optimum correction limit policy by incorporating sampling cost. It is assumed that the sampling cost is a linear function of the sample size.

Let c be the sampling cost at each state and $C_n(T_n)$ be the expected cost per unit time until the completion of $(n+1)$st corrective action under the limit time T_n. Having taken n observations, if we decide to take another observation, i.e., the $(n+1)$st observation, then $C_{n+1}(T_{n+1})$ is the cost per unit time until the completion of the $(n+2)$nd corrective action under the limit time T_{n+1}.

3.1 Cost Function

Let the length of the nth cycle be the time from the beginning of nth operation to the end of nth corrective action. Then the expected cost at the end of $(n+1)$st cycle, given n observations, is

$$E(C_n) = nc + c_1 \int_0^{T_n} \overline{G}(t|\underset{\sim}{y})dt + c_2 \mu_2 \overline{G}(T_n|\underset{\sim}{y}) \quad . \tag{3.1}$$

The expected time to the end of the $(n+1)$st cycle is

$$E[L_n] = \sum_{i=1}^{n} x_i + \hat{x}_{n+1} + \sum_{i=1}^{n} y_i^* + \int_0^{T_n} \overline{G}(t|\underset{\sim}{y})dt + \mu_2 \overline{G}(T_n|\underset{\sim}{y}) \tag{3.2}$$

The expected cost per unit time at the end of $(n+1)$st cycle is then given by

$$C_n(T_n) = \frac{E(C_n)}{E(L_n)} \tag{3.3}$$

or

$$C_n(T_n) = \frac{nc + c_1 \int_0^{T_n} \bar{G}(t|\underset{\sim}{y})\,dt + c_2\,\mu_2\,\bar{G}(T_n|\underset{\sim}{y})}{\sum_{i=1}^{n} x_i + \hat{x}_{n+1} + \sum_{i=1}^{n} y_i + \int_0^{T_n} \bar{G}(t|\underset{\sim}{y})\,dt + \mu_2\,\bar{G}(T_n|\underset{\sim}{y})} \tag{3.4}$$

If we decide to consider the next cycle, then the cost rate function to the end of (n+2)nd cycle is similarly obtained as

$$C_{n+1}(T_{n+1}) = \frac{(n+1)c + c_1 \int_0^{T_{n+1}} \bar{G}(t|\underset{\sim}{y})\,dt + c_2\,\mu_2\,\bar{G}(T_{n+1}|\underset{\sim}{y})}{\sum_{i=1}^{n} x_i + 2\hat{x}_{n+1} + \sum_{i=1}^{n} y_i + \hat{y}_{n+1} + \int_0^{T_{n+1}} \bar{G}(t|\underset{\sim}{y})\,dt + \mu_2\,\bar{G}(T_{n+1}|\underset{\sim}{y})}\,. \tag{3.5}$$

3.2 Optimum Policy

Our objective is to determine the optimum sample size n^* and the optimum correction limit T_n^* such that

$$C_n(T_n^*) \le C_{n+1}(T_{n+1}^*)\,. \tag{3.6}$$

Given that n observations have been taken, the following steps summarize the procedure of determining these quantities:

(i) Calculate $C_n(T_n^*)$ and $C_{n+1}(T_{n+1}^*)$

(ii) If $C_n(T_n^*) \le C_{n+1}(T_{n+1}^*)$, then stop taking observations and employ n^* and T_n^* as the optimum policy.

(iii) If $C_n(T_n^*) > C_{n+1}(T_n^*)$, take (n+1)st observation, i.e. let $n = n+1$ and go to step (i).

The following theorems are used in determining $C_n(T_n^*)$ and $C_{n+1}(T_{n+1}^*)$. Theorems 3.1 and 3.2 are the special cases of the theorems in Appendices A and B, respectively.

Theorem 3.1

Suppose $c_1 < c_2$ and $A \geq 0$. Then there exists a unique and finite T_n^* satisfying

$$r(T_n|\underline{y})[A + (c_2-c_1)\int_0^{T_n} \overline{G}(t|\underline{y})dt] + (c_2-c_1)\overline{G}(T_n|\underline{y}) = B \quad (3.7)$$

where

$$A = c_2\left\{\sum_{i=1}^{n} x_i + \hat{x}_{n+1} + \sum_{i=1}^{n} y_i\right\} - nc \quad (3.8)$$

$$B = \frac{1}{\mu_2}\left[c_1\left\{\sum_{i=1}^{n} x_i + \hat{x}_{n+1} + \sum_{i=1}^{n} y_i\right\} - nc\right]. \quad (3.9)$$

Also, the associated cost rate function is given by

$$C_n(T_n^*) = \frac{c_1 - c_2\mu_2 r(T_n^*)}{1 - \mu_2 r(T_n^*)}. \quad (3.10)$$

Theorem 3.2

If the conditions of Theorem 3.1 hold, then there exists a finite upper bound $\overline{T}_n$ $(> T_n^*)$ such that

$$r(\overline{T}_n|\underline{y}) = \frac{B}{A + (c_2-c_1)\hat{y}_{n+1}}. \quad (3.11)$$

Theorem 3.3

If both T_n^* and T_{n+1}^* exist, then the following relationship holds:

$$\begin{aligned} & C_n(T_n^*) \underset{(\le)}{>} C_{n+1}(T_{n+1}^*) \\ \iff\ & r(T_n^*) \underset{(\le)}{>} r(T_{n+1}^*) \\ \iff\ & T_n^* \underset{(\le)}{>} T_{n+1}^* \ . \end{aligned} \tag{3.12}$$

Relationships given in Theorem 3.3 are explained in Appendix C.

3.3 Numerical Example

In this example we use simulated data to illustrate the determination of n^* and T_n^*. Simulated values of x_n and y_n for various n are given in Table 3.1. Suppose the values of various quantities are as follows:

$$\begin{array}{lll} c_1 = 8000 & c_2 = 9000 & c = 40 \\ \alpha_1 = .8 & \beta_1 = 1 & \\ \alpha_2 = 0 & \beta_2 = 0 & \\ \mu_2 = 0.7 & & \end{array}$$

Then the values of $\hat{x}_{n+1}$, $\hat{y}_{n+1}$, T_n^*, $C_n(T_n^*)$, T_{n+1}^* and $C_{n+1}(T_{n+1}^*)$ are obtained from the above expressions and are given in Table 3.2. From this table we see that for $n = 11$, $C_{11}(T_{11}^*) = 21.74$ and $C_{12}(T_{12}^*) = 23.63$ so that $C_{11}(T_{11}^*) < C_{12}(T_{12}^*)$. Therefore, the optimum policy is $n^* = 11$ and $T_n^* = 0.09$.

TABLE 3.1

Simulated Values of x_n and y_n

n	x_n(hr)	y_n(hr)	n	x_n(hr)	y_n(hr)
1	32.25	1.69	10	3.72	0.15
2	34.77	0.12	11	50.85	0.07
3	63.92	0.23	12	64.89	0.12
4	21.03	0.41	13	0.76	0.29
5	39.42	0.20	14	87.45	1.33
6	9.97	0.37	15	64.12	0.77
7	3.69	0.22	16	30.98	1.37
8	2.42	1.75	17	127.05	1.39
9	10.71	3.00	18	85.54	0.21

TABLE 3.2

Calculations for the Optimum Policy

n	$\hat{x}_{n+1}$	$\hat{y}_{n+1}$	T^*_n	$C_n(T^*_n)$	T^*_{n+1}	$C_{n+1}(T^*_{n+1})$
2	67.01	1.56	0	46.73	0	37.72
3	65.47	1.09	0	32.24	0	32.05
4	50.65	0.91	0.33	30.9	0.33	27.06
5	47.85	0.76	0.92	24.44	0.92	23.74
6	40.27	0.69	0.12	23.00	0.13	20.55
7	34.17	0.62	0.19	21.43	0.19	19.12
8	29.64	0.77	0.95	25.63	0.95	23.25
9	27.27	1.02	0	26.21	0	24.93
10	24.66	0.93	0	26.23	0	24.32
11	27.27	0.85	0.09	21.74	0.09	23.63

4. CONCLUDING REMARKS

In this report we have presented two models for the determination of Bayesian software correction limit policies under the assumption of exponential error occurrence times. For the first model we assume that the sampling cost is negligible while for the second model, such cost is incorporated. Procedures for determining the optimum policy were described and illustrated via numerical examples.

APPENDIX A

Theorem A-1. Let $C^*(T)$ be a cost function given by

$$C^*(T) = \frac{a + c_1 \int_0^T \bar{G}(t|\underset{\sim}{y})dt + c_2 \mu_2 \bar{G}(T|\underset{\sim}{y})}{b + \int_0^T \bar{G}(t|\underset{\sim}{y})dt + \mu_2 \bar{G}(T|\underset{\sim}{y})} \tag{A-1}$$

where a and b are constants and let the following conditions hold

$$c_1 < c_2 \quad \text{and} \quad A \geq 0$$

where

$$A = c_2 b - a .$$

Then there exists a finite and unique T^* which satisfies

$$q(T) \equiv r(T|\underset{\sim}{y})\{A + (c_2 - c_1)\int_0^T \bar{G}(t|\underset{\sim}{y})dt\} + (c_2 - c_1)\bar{G}(T|\underset{\sim}{y}) = B \tag{A-2}$$

where

$$B = \frac{1}{\mu_2}(c_1 b - a) .$$

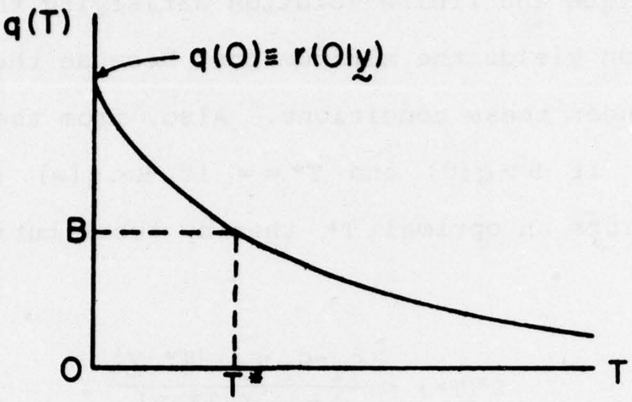
Proof: The solution of the equation $\frac{dC^*(T)}{dT} = 0$ can be expressed in terms of $r(T|\underset{\sim}{y})$ and is given by equation (A-2).

Notice that the repair rate $r(T|\underset{\sim}{y})$ is monotonely decreasing with T. It can be easily shown that the LHS of the above equation, $q(T)$ is also monotonically decreasing with T under the conditions $c_1 < c_2$ and $A \geq 0$. Therefore, if $q(0) > B > q(\infty) = 0$ then there exists a unique and finite solution satisfying the above equation. This solution yields the minimum cost because the cost function is convex under these conditions. Also, from the monotonicity of $q(T)$, $T^* = 0$ if $B > q(0)$ and $T^* = \infty$ if $B < q(\infty)$ (see Figure A-1). If there exists an optimal T^* then by substituting T^* into $C^*(T)$ we get

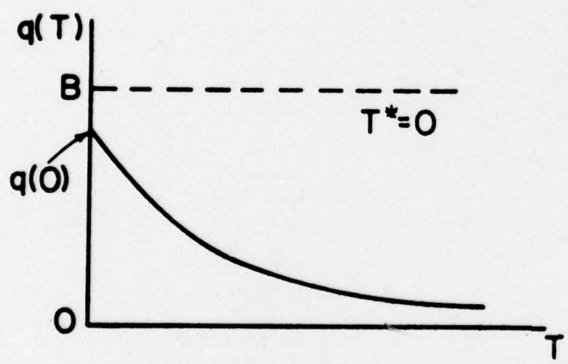
$$C^*(T^*) = \frac{c_1 - c_2 \mu_2 r(T^* | \underset{\sim}{y})}{1 - \mu_2 r(T^* | \underset{\sim}{y})} . \tag{A-3}$$

This completes the proof.

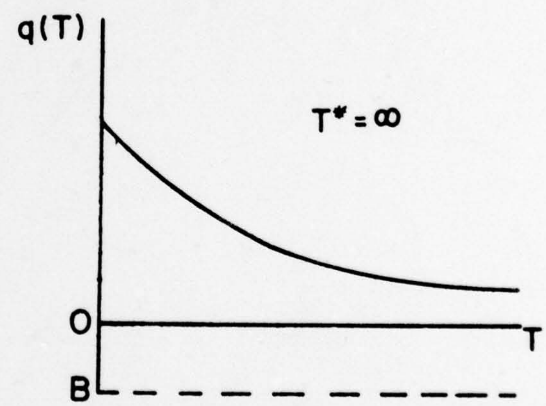
Note that Theorems 2.1 and 2.3 are the special cases of this general theorem.



(a) $q(0) > B > q(\infty) \equiv 0$



(b) $q(0) < B$



(c) $q(\infty) > B$

FIGURE A-1 Plots of $q(T)$ vs T for Various Cases

APPENDIX B

Theorem B-1. *If the conditions of Theorem A-1 hold, then there exists a finite upper bound* $\bar{T}$ $(>T^*)$ *such that*

$$r(\bar{T}|y) = \frac{B}{A + (C_2 - C_1)\,\hat{y}_{n+1}} \qquad \text{(B-1)}$$

Proof of Theorem B-1. We again use the general form to show that there exists a unique and finite upper limit $\bar{T}$ of T^* in case of the existence of T^*, i.e. under the conditions $c_1 < c_2$, $A \geq 0$ and $q(0) > B > q(\infty)$. Since the repair rate $r(T|\underset{\sim}{y})$ is monotonically decreasing with T, we have

$$r(0|\underset{\sim}{y}) > r(T|\underset{\sim}{y}) \quad \text{for} \quad T > 0$$

or

$$r(0|\underset{\sim}{y}) \cdot \bar{G}(T|\underset{\sim}{y}) > g(T|\underset{\sim}{y}) .$$

Integrating both sides over the range of T we get

$$\hat{y}_{n+1}\, r(0|\underset{\sim}{y}) > 1 \qquad \text{(B-2)}$$

where

$$\hat{y}_{n+1} \equiv \int_0^\infty \bar{G}(T|\underset{\sim}{y})\,dt$$

is the posterior mean. Now define the following function.

$$M(T) = r(T|\underset{\sim}{y})\{A + (c_2 - c_1)\hat{y}_{n+1}\} - q(T) . \qquad \text{(B-3)}$$

Then, from B-2 we note that $M(T)$ is monotonically decreasing in T with

$$M(\infty) = 0 \qquad \text{(B-4)}$$

and

$$M(0) = (c_2-c_1)\hat{y}_{n+1}\, r(0|\underset{\sim}{y}) - (c_2-c_1)$$

$$= (c_2-c_1)\{\hat{y}_{n+1}\, r(0|\underset{\sim}{y}) - 1\} > 0 \tag{B-5}$$

Therefore,

$$M(T) > 0 \quad \text{for} \quad T > 0$$

or

$$r(T|\underset{\sim}{y})\{A + (c_2-c_1)\hat{y}_{n+1}\} > q(T) \ . \tag{B-6}$$

Hence, if there exists a T^* such that

$$q(T^*) = B \tag{B-7}$$

then there also exists a unique and finite root $\bar{T}$ satisfying

$$r(\bar{T}|\underset{\sim}{y})\{A + (c_2-c_1)\hat{y}_{n+1}\} = B$$

or

$$r(\bar{T}|\underset{\sim}{y}) = \frac{B}{A + (c_2-c_1)\hat{y}_{n+1}} \ . \tag{B-8}$$

It is easily seen that $\bar{T} > T^*$ (see Figure B-1).

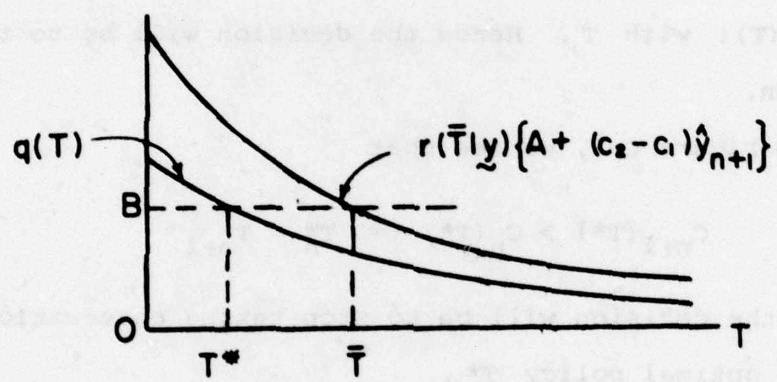


FIGURE B-1 Plot of q(T) vs T

APPENDIX C

Relationship between $C(T_n^*)$ and T_n^*

If $c_1 < c_2$, then $C(T_n^*)\downarrow$ with $r(T_n^*)$. This is seen to be true from equation (2.25). Now, from Figure C-1, we see that

$$C_n(T^*) > C_{n+1}(T^*) \Longleftrightarrow T_n^* > T_{n+1}^*$$

because $r(T)\downarrow$ with T. Hence the decision will be to take another observation.

From Figure C-2, we see that

$$C_{n+1}(T^*) > C_n(T^*) \Longleftrightarrow T_n^* < T_{n+1}^*$$

and hence the decision will be to stop taking observations and to employ the optimal policy T_n^*.

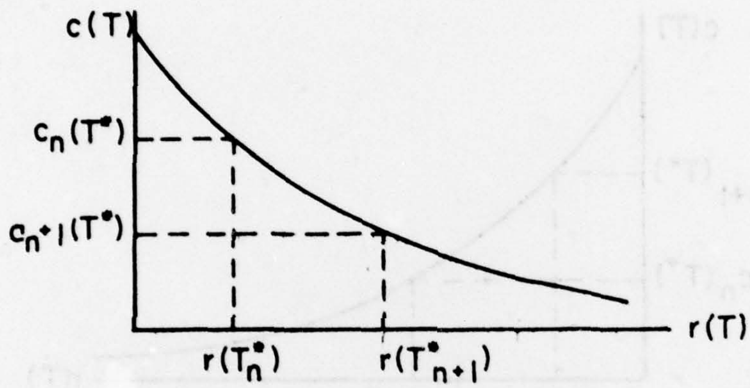


FIGURE C-1 Plot of C(T) versus r(T)

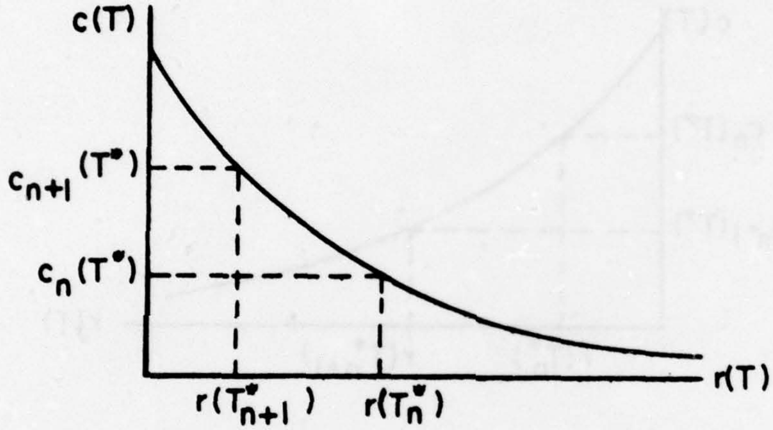


FIGURE C-2 Plot of C(T) vs r(T)

REFERENCES

[1] Barlow, R. E. and F. Proschan, Statistical Theory of Reliability and Life Testing. Holt, Reinhart and Winston, N.Y., 1974.

[2] Boehm, B. W., "Software and Its Impact: A Quantitative Assessment," Datamation, pp. 48-59, 1973.

[3] Box, G. E. P. and G. C. Tiao, Bayesian Inference in Statistical Analysis. Reading, MA: Addison-Wesley Publishing Co., 1973.

[4] Drinkwater, R. W. and H. A. J. Hastings, "An Economic Replacement Model," Oper. Res. Q., Vol. 18, pp. 121-138, 1967.

[5] Hastings, N. J., "The Repair Limit Replacement Method," Oper. Res. Q., Vol. 20, pp. 337-349, 1969.

[6] Hecht, H., "Can Software Benefit from Hardware Experience?" Proceedings 1975 Annual R & M Symposium, pp. 480-486, 1975.

[7] Lambe, T. A., "The Decision to Repair or Scrap a Machine," Opl. Res. Q., Vol. 25, pp. 99-110, 1974.

[8] Nakagawa, T. and S. Osaki, "The Repair Limit Replacement Policies," Oper. Res. Q., Vol. 25, pp. 311-317, 1974.

[9] Okumoto, K. and S. Osaki, "Repair Limit Replacement Policies with Lead Time," Zeitschrift fur OR, Vol.20, pp.133-142, 1976

[10] Osaki, S. and K. Okumoto, "Repair Limit Suspension Policies for a Two-Unit Standby Redundant System with Two Phase Repairs," Microelectronics and Reliability, Vol. 16, pp. 41-45, 1977.